



PROCEEDINGS OF THE ELEVENTH ANNUAL ACQUISITION RESEARCH SYMPOSIUM

THURSDAY SESSIONS VOLUME II

The Efficacy of the Government's Use of Past Performance Information: An Exploratory Study

Tim Hawkins
Western Kentucky University

Published April 30, 2014

Approved for public release; distribution is unlimited.

Prepared for the Naval Postgraduate School, Monterey, CA 93943.



ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 30 APR 2014		2. REPORT TYPE		3. DATES COVERED 00-00-2014 to 00-00-2014	
4. TITLE AND SUBTITLE The Efficacy of the Government's Use of Past Performance Information: An Exploratory Study				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Western Kentucky University,1906 College Heights Blvd,Bowling Green,KY,42101				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Since its inception via the Federal Acquisition Streamlining Act of 1994, contractor past performance is intended to be an important evaluation criterion in federal source selections. In order to reduce performance uncertainty, procurement officials must record contractor performance evaluations in a central database. However, reports of ubiquitous problems raise questions of the integrity of ratings and the utility of the evaluations. From a literature review, several factors affecting the efficacy of past performance evaluations are identified. These factors are combined in a comprehensive conceptual model explaining past performance efficacy. Exploratory, qualitative data preliminarily confirms the hypotheses. Key antecedents include the following: rating justification quality; contractor surveillance; multirater dissonance; perceived accuracy; evaluator role overload; fear of supplier dispute perceived fairness; sufficiency of requirement definition; evaluator turnover; relationship quality; and buyer???supplier communication frequency, bi-directionality, and formality. From these findings, important managerial and theoretical implications are drawn and future research directions are identified.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 33	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

The research presented in this report was supported by the Acquisition Research Program of the Graduate School of Business & Public Policy at the Naval Postgraduate School.

To request defense acquisition research, to become a research sponsor, or to print additional copies of reports, please contact any of the staff listed on the Acquisition Research Program website (www.acquisitionresearch.net).



ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL

Panel 15. Advances in Contract Management

Thursday, May 15, 2014	
11:15 a.m. – 12:45 p.m.	<p>Chair: BG Stephen B. Leisenring, USA, Deputy Director of Contracting, U.S. Army Corps of Engineers</p> <p><i>The Efficacy of the Government's Use of Past Performance Information: An Exploratory Study</i> Tim Hawkins, Western Kentucky University</p> <p><i>Contracting for Reform: The Challenges of Procuring Security Training and Advisory Services in Fragile Environments</i> Nicholas Armstrong, Syracuse University David Van Slyke, Syracuse University</p> <p><i>Price Analysis on Commercial Items Purchases Within the Department of Defense</i> Ralucca Gera, Naval Postgraduate School Janie Maddox, Naval Postgraduate School</p>



The Efficacy of the Government's Use of Past Performance Information: An Exploratory Study

Timothy G. Hawkins—Lt Col, USAF, PhD, is an assistant professor at Western Kentucky University in Bowling Green, Kentucky. Lt Col Hawkins has published articles on opportunism in buyer–supplier relationships and on electronic reverse auctions in scholarly publications such as the *Journal of Supply Chain Management*, *Journal of Business Research*, *Journal of Business Ethics*, *Industrial Marketing Management*, *Supply Chain Management: An International Journal*, the *International Journal of Logistics Management*, the *Journal of Marketing Channels*, *Air Force Journal of Logistics*, and the *Journal of Public Procurement*. His current research interests include electronic reverse auctions, procurement ethics, buyer–supplier relationships, strategic sourcing and services procurement. [timothy.hawkins@wku.edu]

Abstract

Since its inception via the Federal Acquisition Streamlining Act of 1994, contractor past performance is intended to be an important evaluation criterion in federal source selections. In order to reduce performance uncertainty, procurement officials must record contractor performance evaluations in a central database. However, reports of ubiquitous problems raise questions of the integrity of ratings and the utility of the evaluations. From a literature review, several factors affecting the efficacy of past performance evaluations are identified. These factors are combined in a comprehensive conceptual model explaining past performance efficacy. Exploratory, qualitative data preliminarily confirms the hypotheses. Key antecedents include the following: rating justification quality; contractor surveillance; multi-rater dissonance; perceived accuracy; evaluator role overload; fear of supplier dispute; perceived fairness; sufficiency of requirement definition; evaluator turnover; relationship quality; and buyer–supplier communication frequency, bi-directionality, and formality. From these findings, important managerial and theoretical implications are drawn and future research directions are identified.

Introduction

Industrial buyers labor to avoid the deleterious effects of the laws of agency. In industrial buying, the supplier serves as an agent to the principal (buying organization). Substantial effort is dedicated to avoid adverse selection and moral hazard. *Adverse selection* encompasses the risk of selecting an incapable supplier that otherwise misrepresents itself as capable, while *moral hazard* is the vulnerability to acts of supplier opportunism (Eisenhardt, 1989)—behavior that is self-interest seeking with guile (Williamson, 1975). For example, supplier opportunism could include shirking quality, obfuscating the truth, withholding information, lying, cheating, and breaching contract terms (Wathne & Heide, 2000).

In their buying efforts, government agencies incur significant transaction costs attempting to write all-inclusive contracts and to monitor contractor performance in order to thwart supplier opportunism. These costs of contracting are substantial given the magnitude of contracted goods and services. In fiscal year (FY) 2010, the federal government awarded more than 5.9 million contract actions worth over \$538 billion (Federal Procurement Data System—Next Generation [FPDS–NG, n.d.]). More transaction costs are incurred attempting to mitigate information asymmetries, thereby avoiding adverse selection, by requiring that past performance be an evaluation criterion for contract award. The logic is that by better informing source selection decisions, better best value selections will occur. Integrally related is the contractor's performance; if performance levels are assessed and recorded, and if this information is available to future source selection teams, conventional wisdom holds that contractors will work harder to ensure satisfactory (or better) performance.



In U.S. federal government contracting, agencies are required to consider past performance information as an evaluation factor in formal source selections. Necessarily, then, agencies must collect and report contractor past performance information from certain government contracts. However, there are many concerns that the past performance evaluations/ratings are not properly, timely, or accurately completed. From 2007 to 2010, overdue assessments grew from 5.3% to 10.1% of total assessments required (Contractor Performance Assessment Reporting System [CPARS] Metrics, n.d.). In 2009, the Government Accountability Office (GAO) estimated that only 31% of contract actions requiring CPARS reporting had completed reports. Reports often lack sufficient information to support ratings (e.g., how the contractor met, exceeded, or failed to meet requirements) necessary to withstand a legal challenge, or do not include a rating for all performance areas (Office of Federal Procurement Policy [OFPP], 2011). Additionally, throughout the rating process, raters often inflate ratings in order to avoid conflict with the contractor (GAO, 2009).

Unreliable or inaccurate past performance assessments can harm contractors' reputations and can bias source selections resulting in adverse selection. If past performance information is not reliable, and if contracting officers and evaluators do not (or cannot) use the information to discriminate between competitive proposals (Kelman, 2010), the effort of collecting and reporting the past performance information is squandered. Likewise, the effort of evaluating and documenting inaccurate past performance information during source selections is wasted. Federal contract managers are already overworked (GAO, 2009) and understaffed (GAO, 2001); therefore, continuing to consume time on a fruitless task would be futile.

While the GAO (2009) suggested that assessments and ratings are inflated, the degree of inflation is unknown. Evidence suggests that the magnitude of distortion is high—so much that contracting officers, evaluators, and source selection authorities rarely use past performance information as a meaningful discriminator between proposals. In order to determine whether this seemingly vacated faith is warranted, the degree of distortion needs to be assessed. The extent of distortion will tell us whether the reporting system and policy need to be abandoned, adjusted, or left intact.

The purpose of the research, therefore, is to explore the efficacy of the government's current use of past performance information. The intent is to diagnose alleged weaknesses and to explore potential improvements. The following research questions are addressed:

1. Are past performance reports useful? How so, or why not?
2. In the cases of multiple evaluators on a single contract action, do past performance evaluations/ratings deviate among evaluators, and, if so, why?
3. Why do reviewing officials change the ratings of the evaluator (assessing official)?
4. To what extent do past performance evaluations/ratings captured in federal databases influence source selection decisions?
5. Why do past performance evaluations/ratings lack sufficient justification/supporting information?
6. Why are past performance evaluations sometimes inaccurate?

The answers to these six questions should help diagnose the efficacy of the government's current collection and use of past performance information. The remainder of this paper is organized in the following manner. First, Figure 1 displays the conceptual framework and proposed hypotheses. The theoretical underpinnings of this model will not be



discussed here. Next, the study presents the research design and methodology. Lastly, discussion, limitations, implications, and future research directions are offered.

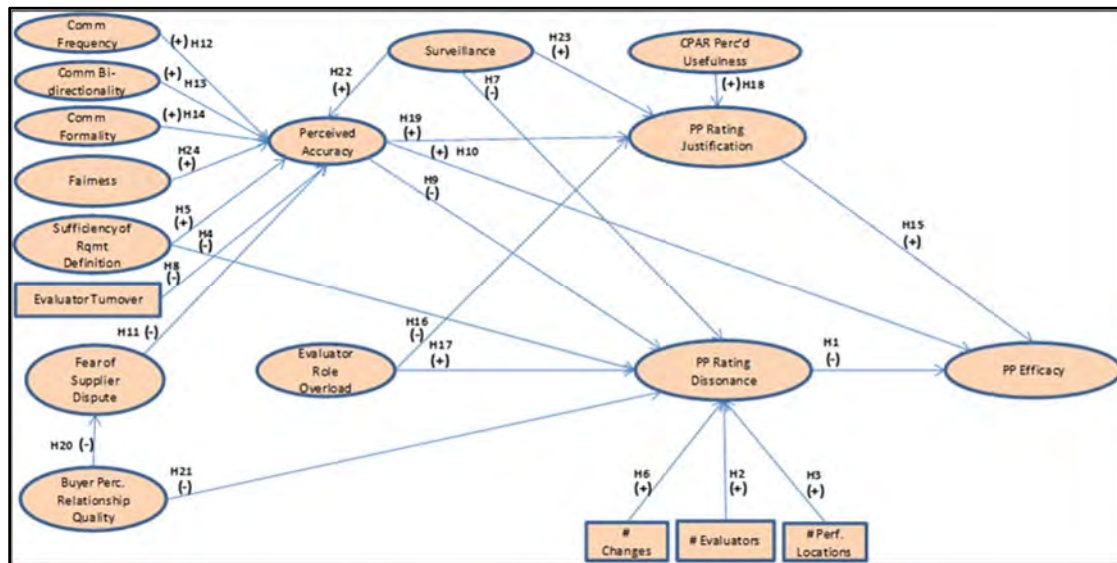


Figure 1. Conceptual Model

Note. Ovals represent latent constructs; rectangles represent objective measures.

Methodology

Qualitative Data Analysis

This research used a qualitative methodology to examine the efficacy of past performance evaluations. According to Yin (2009), a qualitative methodology is appropriate when three conditions exist: (1) The type of research question is exploratory in nature and takes the form of a “why” question, (2) the researcher has no control of the behavioral events being researched (i.e., cannot manipulate behaviors then measure results as in a controlled experiment), and (3) the focus is on contemporary events (p. 8). Furthermore, case study research is particularly useful when researchers need to provide insight and depth to a unique phenomenon (Ellram, 1996).

Data Collection

The interview protocol was developed based on a review of archival CPARs, the literature surrounding supplier performance evaluation, and discussions with academic experts and participants involved with past performance evaluations and source selections. In all, eight interviews were conducted. The interviews lasted between 38 and 67 minutes (mean 51 minutes). Each interview was recorded, then transcribed. Transcripts were then sent to informants for an accuracy check, thereby enhancing construct validity (Flint, Woodruff, & Gardial, 2002; Yin, 2009). Transcripts averaged 18 pages and 7,394 words in length.

Data Analysis

The analysis process began by identifying constructs, defining those constructs, and then positing relationships between them (Patrick Van Eecke, 2006). Each interview was examined to identify themes and then tested to determine whether these themes remained consistent in subsequent interviews or in reexaminations of previous interviews. The

participant interviews continued over a period of eight weeks. Initial coding led to new interviews with new participants to gain clarification and validation.

Sample

The sample of informants (Table 1) was drawn from the researcher's personal contacts within one military service. Military and civil service employees who routinely evaluate contractor performance and enter these evaluations into the CPARS participated. These experts represented two industries that account for a large portion of the federal government's portfolio of contract spending, aerospace and information technology (IT). Experience in evaluating contractor performance ranged from two to 28 years, and there was a similar wide range of the number of past performance evaluations experienced (1–50). Most informants were program managers since they often assume responsibility for reporting past performance evaluations. One contracting officer with extensive experience in CPARS, both in reporting and evaluating CPARS during source selections, was included.

Table 1. Informant Demographics

Informant	Civilian/ Military	Industry	Years Experience	Role	Past Performance Experience (Number of Evaluations)
1	Civilian	Aerospace	28	Contracting Officer	50+
2	Military	Aerospace	7	Program Manager	10
3	Civilian	IT	4	Program Manager	11
4	Civilian	IT	10	Program Manager	7
5	Military	IT	10	Program Manager	5
6	Military	IT	9	Program Manager	15
7	Military	IT	2	Program Manager	1
8	Military	IT	18	Program Manager	10

Results

The result of each research question is discussed in sequential order followed by excerpts from interview informants. The meanings of the excerpts are then discussed and related back to the hypothesized relationships represented in the conceptual model (Figure 1).

1. Are past performance reports useful? How so, or why not?

To examine whether past performance evaluations are seen as useful, we adopted the commonly touted utilities of past performance information to (1) reduce performance risk in future source selections, thereby reducing contractor performance uncertainty, and (2) motivate contractor performance. Of the seven informants commenting on this question, the results were mixed; three agreed that past performance evaluations reduce performance risk, while four disagreed.

I think it could be effective at mitigating a risk if the requirements that you are looking at match up with the [inaudible] past performance evaluations that you are comparing them to.



This informant qualified a past performance evaluation as useful if it is relevant to the requirement under consideration during source selection. For source selections, relevance is a requisite criteria of past performance evaluations.

It was a lot of fluff and I am afraid that unless everyone is really working these things to really make an impactful statement that they probably aren't worth a whole lot if you have a lot of ones that just are fluffy.

Because you can't adequately make an assessment of a contractor's potential to perform on the future based on a ball of fluff.

These separate informants complained that a lack of specific details hindered the utility of past performance evaluations. In other words, a lack of details can render the evaluation useless. Additionally, a lack of details can render the judgment of relevance difficult.

I know that it is going to be watered down kind of like the [enlisted performance report/officer performance report] because there is so much pressure that the contractor puts back on the government for wording intricacies. Overall, I think I would have to question the overall overarching fairness of the process just because just like the [enlisted performance report/officer performance report] system, particularly the [officer performance report] system you question how much reality you are getting out of this if you are not seeing all of these support that goes behind the ratings. That is why I would have to say overall I would question it.

Drawing a parallel to Air Force military personnel performance appraisals, this informant essentially commented that the past performance evaluations are inflated so as to not harm the contractor. This comment suggests support for H10, that the efficacy of a past performance evaluation could be hindered by an inaccurate (i.e., inflated) report. The next part of this comment (i.e., "pressure") suggests a fear of a contractor's dispute of the narrative assessments and/or ratings. In the context of the conversation, this testimony suggests support for H11, that fear of a contractor dispute may decrease the accuracy of the evaluation (i.e., rating inflation). The testimony also suggests that detailed rating justifications are needed in order to extract value (i.e., usefulness) from the past performance assessment, thus, supporting H15.

One informant commented,

I think in concept it is not that bad. In application, it varies a lot and it is hard to get a total—the whole CPAR system is fair or not fair. I've seen it be fair in some places and I have seen it not be fair in some places. I have seen just a very mixed bag in a lot of places. I have seen some places and people running around with their hair on fire and it is just a task to do and they slam something out at the last second.

This testimony infers that (1) there is variance in how past performance reports are accomplished and their quality, and (2) some assessing officials (raters) do not value the report—calling into question its utility.

Of the six informants commenting on the second part of this question, four agreed that past performance evaluations motivate contractor performance, while one informant disagreed:

I think [a past performance evaluation] does motivate contractors to a certain extent.



It can be a great tool for the PM to use to motivate the contractor. I see its effectiveness on that end more so than on a source selection, if you will.

Researcher: So do you think—at least in your experience in those types of programs, do the CPARS tend to motivate contractors to perform?

Informant: I would say very minimally. It became more of an exercise of they did what they do. Then you back into these ratings and then we had a person come along different up the food chain who would review those before they went out and had different standards for what the different colors meant.

While results were mixed as to whether past performance evaluations reduce performance risk for a future contract, most informants agreed that the evaluations motivate contractors to perform better.

2. In the cases of multiple evaluators on a single contract action, do past performance evaluations/ratings deviate among evaluators, and, if so, why?

Of the five informants commenting on this question, each affirmed cases in which a contract involved multiple different performance evaluators (H2). One informant commented,

Sometimes there was some real consternation, and sometimes they actually went outside the program team and went up to higher management to get it resolved.

The informants offered a variety of explanations for differences in assessments. Three informants mentioned different expectations of contractor performance and poor requirements definition as culprits, confirming H4 and H6 (number of changes). Two informants attributed incongruent past performance evaluations to insufficient monitoring of the contractor. This supports H7. Two informants mentioned that the different government performance evaluators had different experiences, suggesting that individual differences may exist. Two informants mentioned different locations of the contracting officer's representative, indicating that performance may differ at different physical sites, supporting H3. Two informants also agreed that work overload precludes performance evaluators from fulfilling their duties to evaluate and document contractor performance, supporting H16 and H17.

Informant: You have only got so many resources, and I see a number of program offices that they are doing so many things they are driving ahead of their headlights.

Researcher: So workload is an issue?

Informant: Workload is a definite.

The following four additional reasons for dissonance among performance evaluators included a lack of facts of performance levels (H9), fear of supplier dispute of the ratings, rater revenge, and differences in standards for ratings across evaluators (H4).

3. Why do reviewing officials change the ratings of the evaluator/assessing official)?

When inquiring whether reviewing officials change ratings and/or narrative statements made by performance evaluators, the results were mixed. There appears to be plenty of opportunity for changes since several layers of management review a CPAR, as evidenced by one informant.



From here and my boss looks at it and he is actually the program manager, [inaudible]. Then we get past them to [inaudible] one, two, three—I would say three. Three layers. If you include the contractor who eventually has a chance to look at it, that is probably a fourth layer.

Three informants confirmed the practice, while three had no experience with changed evaluations. For those experiencing changed ratings, reasons cited included a lack of facts of contractor performance and government responsibility for contractor nonperformance.

Researcher: You would see narrative and ratings get changed?

Informant: In some cases.

Researcher: They got changed outside of what was truly accurate or earned or deserved?

Informant: Many—in my opinion, many of the ratings for a long time could have been a lot lower if government had its act together and adequately supported and communicated with the contractor.

This exchange attributes changed ratings to the government's failure to observe or document contractor performance. The informant also mentioned a failure to communicate with the contractor.

When discussing a fear of a contractor's dispute (i.e., a potential claim) of a past performance assessment, one offeror alluded hypothetically to diminished value of the CPAR in achieving its intended objectives. The informant then likened a change in CPARS reporting presumably to a change in a source selection rating of past performance upon being disputed (again, presumably via a bid protest). This testimony offers some evidence that a fear of a contractor's dispute is germane to the accuracy of past performance evaluations (H11), and associated this fear to diminished past performance efficacy (H10).

Let's say if this gets to a legal—if we get to the point in a CPAR—and how we do a CPARS or contractor assessments—to where we are concerned and it becomes a legal fear, then I think that the value of them will disappear and there will be no value in there. I say that from experience in a couple of situations. One of the past performance teams I led, ultimately there was a protest, and we successfully defended against the protest. The protest was denied. But, it was because we had clearly worked with legal ahead with our sections L and M in the RFP and we stuck by that methodology and we documented our methodology. And like the guy that came behind me—I deployed for a little while—and the guy that was leading the experience team took over the past performance team—he ended up spending about three days on the stand—of significant grilling. But because we had well-documented processes and we had not deviated from our section L—how we told them we were going to evaluate them—and we could substantiate them in the thing, I had no fear that we were going to [inaudible]. So we had right on our side, so I had no fear of standing by what we had done. In another situation I was involved in where—in process not CPARS, but very similar, to where a lot of information got watered down and changed once it became a legal matter and legal process.

4. To what extent do past performance evaluations/ratings captured in federal databases influence source selection decisions?



This question resembles the first part of the first research question. Question 1 inquired whether past performance evaluations are useful to reduce performance risk on future contracts. These results are less mixed, with most informants believing that past performance evaluations do not influence source selection decisions (i.e., winner determinations). One informant reported no influence. Three informants reported little influence. One informant reported some influence, and one informant reported great influence.

5. Why do past performance evaluations/ratings lack sufficient justification/supporting information?

Several informants confirmed that often past performance evaluations lack sufficient justifications for ratings and narrative assessments. In explaining why past performance evaluations lack sufficient justifications, several informants identified poor documentation of contractor performance. Poor documentation of facts could result from excess workload or a lack of contractor surveillance. Thus, support is found for H19 and H23. Two informants also identified evaluator turnover as a culprit:

And there is a wide variety within the system, in my experience. So you get—and you find that out by calling back to the PMs that you can get ahold of, if they are still there. The older the CPARS are, obviously it is harder to find the people, and you clarify the information you are reading from a past performance perspective.

The informant, here, referred to a high variance in quality of past performance assessments, so much so that in many cases, phone calls back to the program manager are necessary in order to validate and understand the contractor's performance. However, this understanding is hindered by a turnover of personnel who generated the CPAR. Another informant highlighted the effect of his turnover on a CPAR:

I was working on another project completely different from this and couldn't even spell CPAR. I mean I didn't really know what it was and all of a sudden I was made the program manager for a certain—for a program—and it came to, okay, it is time to do their CPAR. I wasn't even—it was like, okay, I worked with the contractor and you know worked with the contractor to come up with what she wanted in the CPAR. Okay? At that point I was like, okay, I will write something up and send it over to them, and if it is okay with them, then we will send it forward and that was probably—I know now that is okay, you get input from them but then it is actually you writing it and then you don't have to necessarily—you don't have to always agree with what the contractor thinks they did. I mean sometimes you can think differently. So my first one was—and I don't even remember what the ratings were—I really don't, but I know that first one, that was probably—I am not going to say it was wrong, but I am going to say it was—I couldn't have backed up some of the stuff that was in there because I wasn't working with the contractor.

In this case, since the informant had no experience with CPARS reporting and since, due to his recent turnover, he was not cognizant of the contractor's performance, he essentially let the contractor write its own CPAR. Thus, support is also found for H8—that evaluator turnover diminishes the accuracy of CPARS.

6. Why are past performance evaluations sometimes inaccurate?

Informants unanimously and strongly agreed that past performance evaluations, too often, are inaccurate. Many explanations were provided by the seven informants responding



to this question. Informants mentioned the following factors affecting accuracy: halo effect (unwillingness to taint a contractor's record since it could effectively lock them out of future awards), lack of facts surrounding contractor performance, inflated ratings, performance evaluator turnover (H8), differing definitions of performance standards (H5), poor requirements definition (H5), poor oversight of contractors (H22), and the disregarding of some deficiency reports.

That is very hard to get an under satisfactory from what I have seen.

Many—in my opinion, many of the ratings for a long time could have been a lot lower if government had its act together and adequately supported and communicated with the contractor.

Some services tend to not put much negative information in there in my experience. At least the ones I have read. Some of them are written more like a performance report where it's bad to say anything negative. I think that—if that is the approach that people take, then you would take then the system has little value.

These testimonies of separate informants confirm inflated ratings and the halo effect, which compromises accuracy. One reason underlying the inflated rating—to protect the contractor from a permanent scar—could be attributed to a concern for fairness, supporting H24. Another reason is the government's failure to observe and document contractor performance (H22).

Researcher: To what extent do you guys worry about a dispute from a contractor or rebuttal?

Informant: I think the way that you address that or minimize the chance of that happening, you know, along the same lines of what these guys had said. Number one, shouldn't be any surprises on a CPAR. CPAR should not be the first time that the contractor hears about an issue. Then number two, being objective on a CPAR. If you can trace it back to your requirements or PWS and you have an objective affirmation on there, I think that reduces the chance of that happening a lot.

This quote suggests that, consistent with H11, the fear of a contractor's dispute of the ratings or narrative assessments influences performance evaluators to collect and document supporting facts. These fact-based evaluations should improve the accuracy of the past performance evaluation.

There were other things that were like, well, they didn't perform as well as we wanted them to, but we couldn't ding them on it because nowhere in the contract did it specifically say this is your standard and this is where you have to meet it or exceed it.

Researcher: Does anybody have any experiences with accuracy—you know, issues of accuracy of the CPARS that you could tie back to something like a poorly defined requirement or not the proper amount of oversight or surveillance to the contractor?

Informant: We have seen a few of those things which makes the documentation part harder—or not documentation, but the supporting arguments harder, when you say, "Okay, well their requirement is this." Well, how do you meet that because you can't even define that?



These quotes suggest that sometimes performance requirements are not sufficiently defined in order to collect facts and compare them to contractual requirements. Thus, support is found for H5.

The division leadership and this particular organization has pushed down a culture that lends itself to that evidence in writing CPARS. You know the division staff pushes it down to the branch level, and the branch reviewers push that down too. So that is the first thing they look for when they are reviewing the write ups is, okay, now give me the four examples. You know if you have gone above and beyond, give me an example of that. If you have a lack of communication, give me examples of that. So that is a culture that has been pushed down to this division and that is the expectation that is displayed. The reason for that is we don't want to go down the road for dispute. That is our defense mechanism in this particular division.

So we work hard in this division to have the evidence within the CPAR so it doesn't get disputed down the road if we run into issues.

This testimony confirms a fear of a supplier dispute, and demonstrates that this fear influences performance evaluators to bolster the justifications of their past performance ratings and narratives.

Yes, when I was [in] the last program office that I was in, we had our support contractor, and we were meeting with that contractor virtually through email and through telephone conversation multiple times a week and constantly giving feedback. So when it was CPARS time, there were no surprises. Actually it didn't even get disputed, and we had a couple of areas where we had a few markdowns and we had the data, and that is the important thing in writing is the data to back it up. You know, dates and documented evidence, if you will, [inaudible] come to that for an area that they may have been lacking in. So it wasn't a surprise, just to my [inaudible] it was not a surprise for the [inaudible] contractor to get the CPAR that they did. It was constant feedback and that was just in the way of the working relationship.

This exchange suggests an association between the buyer–supplier relationship and the quality and frequency of communications. The informant mentioned *no surprises* and *no disputes* from the contractor due to the communication. This insinuates that the evaluations were accurate and that there is, therefore, little concern for a supplier dispute. Thus, some support appears reasonable that, consistent with H20, relationship quality affects a fear of a dispute (which, in turn, affects the accuracy of the past performance evaluation).

Nonetheless, there appears to be significant variance in the rigor, frequency, quality, and amount of performance feedback across contracts. These features of communication vary by individual program managers, contract managers, CORs, or end users. This variance lends credence to H12, H13, and H14, which posit relationships between features of communication and past performance evaluation accuracy.

In general, qualitative interview data largely supports the conceptual model lending content validity. Informants did not specifically identify associations between past performance efficacy and evaluator dissonance (H1). Nor did they explicitly link CPAR usefulness to past performance rating justifications (H18). Most of these relationships are, however, implicit in the conversations. Explicit support discussed above was found for the remaining 21 hypotheses.



Discussion

Since its inception via the Federal Acquisition Streamlining Act of 1994 (Beausoleil, 2010), contractor past performance is intended to be an important evaluation criterion in federal source selections. The purpose was to level the playing field between the government and the contractors to mitigate information asymmetries. With more complete knowledge of contractor performance, agencies can mitigate adverse selection.

However, there are many concerns that the past performance evaluations/ratings are not properly, timely, or accurately completed. Reports often lack sufficient information to support ratings (e.g., how the contractor exceeded or failed to meet requirements) necessary to withstand a legal challenge or do not include a rating for all performance areas (OFPP, 2011). Additionally, throughout the rating process, raters often inflate ratings in order to avoid conflict with the contractor (GAO, 2009).

Unreliable or inaccurate past performance assessments can harm contractors' reputations and can bias source selections, resulting in adverse selection. If past performance information is not reliable, and if contracting officers and evaluators don't use it in discriminating between competitive proposals, the effort of collecting and reporting the past performance information is squandered. Likewise, the effort of evaluating and documenting inaccurate past performance information during source selections is wasted. Evidence suggests that the magnitude of distortion is high—so much that contracting officers, evaluators, and source selection authorities rarely use past performance information as a meaningful discriminator between proposals. In order to determine whether this seemingly vacated faith is warranted, the degree of distortion was examined.

The purpose of the research was to explore the efficacy of the government's current use of past performance information. The intent was to diagnose alleged weaknesses and to explore potential improvements. This research used a qualitative methodology to examine these research questions. From a literature review, a conceptual model of 24 hypotheses was developed. Eight subject matter experts who routinely evaluate contractor performance and enter these evaluations into the CPARS were interviewed to explore the relationships posited in the model. While employing only a limited, qualitative, empirical test of the propositions, the research provides managers with some tentative guidance.

Managerial Implications

This research confirmed much of what has been reported in GAO and OFPP reports. However, the research took the next step to explain why the systemic weaknesses occur (e.g., inflated ratings, poor justifications, etc.). In doing so, several novel causal factors emerged. For example, some main findings centered around the dissonance among multiple performance evaluators on a single contract. Another major finding entailed the accuracy of evaluations and how the characteristics of channel communication play such an important role in accuracy. The findings herein introduce a plethora of implications for acquisition management, discussion of which follows.

First, dissonance across performance evaluators suggests that contractors should pay attention to evaluator dissent and develop strategies to manage each of the buyer's agents' interpretations of its performance. Government acquisition teams and contractors might benefit from discussing during the post-award conference precisely how a situation of dissent among multiple evaluators will be managed. Additionally, since evaluator workload can affect the due diligence applied to performance evaluations, contractors could devise strategies to make the evaluators' jobs less arduous. For example, contractors can, and sometimes do, preempt the CPAR by writing their own versions of evaluations and offer these evaluations to evaluators, program managers, and contracting officers. The



unintended consequence of this practice, however, is the buyer's propensity to apply less effort in its duties to independently monitor and scrutinize performance. Where buyer-supplier trust is high and where contractor performance is high and reliable, this practice of essentially outsourcing performance evaluations poses less risk. Agencies should, however, weigh the conflict of interest posed and set boundaries for this practice since it invites risk of artificially inflated assessments.

The research also offers explanations for dissenting evaluations among multiple performance evaluators. For example, leaders should manage evaluator workload to ensure they have sufficient time to perform their past performance evaluation duties. Manning models should be more precisely developed to account for not only dollars obligated and the number of contracts awarded annually, but other time-consuming tasks such as the quantity of past performance evaluations. This research reveals that, on average, past performance evaluations consume nearly two man-weeks of effort. Leaders should also devise means to ensure that requirements—including measurements of success and precise definitions of CPAR ratings tailored to the requirement—are sufficiently defined prior to solicitation. These definitions should be reviewed at the post-award conference. Where interpretation can vary among evaluators, different expectations of contractor performance can emerge and fester. Likewise, the number of changes should not be excessive since this, too, can result in confusion as to what is required by the contractor, particularly on high-value, complex requirements. Inter-rater dissonance may also be reduced by ensuring that past performance assessments and ratings are more fact-based (i.e., more accurate) since it is difficult to disagree with documented facts. Finally, leaders can reduce dissonance with more proper surveillance of the contractor's work.

A central construct affecting past performance efficacy appears to be the accuracy of the evaluations. Accuracy was found to be affected by many fairly obvious factors that have been discussed in the literature, such as increased surveillance, feedback quality, communication bi-directionality, communication formality, and fear of a supplier dispute (resulting in a halo effect). These results suggest that more surveillance and performance-level measurement should be conducted in order to observe and collect the requisite facts. Thus, requiring activities should develop metrics to assess contractor performance and schedules for measurement.

The results also suggest that past performance reporting is often not a sufficient surrogate for contractor performance management. More frequent, formal, and two-way communication with the contractor is usually required, as affirmed by Steve Kelman's (2010) recommendations to improve past performance information collection and use. Thus, acquisition teams relying on the CPARS system as the sole feedback mechanism may sacrifice accuracy and, in turn, past performance efficacy.

This research highlights the limitation of CPARS and a gap in federal procurement management. There is no single structured IT system and process to systematically collect, store, and synthesize contractor performance information. This is one reason why the government struggles so much to effectively manage service contracts. Supplier performance management systems are common in the for-profit sector. Examples include lasta's SmartSupplier scorecard tool, SAP/Ariba's *Supplier Performance Management* module, and BravoSolution's *Supplier Performance Management* tool. These structured, web-enabled tools could standardize metrics, performance data recording, analysis, and reporting. They also offer dashboard-like scorecards to assess individual contractors and groups of contractors (e.g., by commodity family or by industry). Such a structured tool could alleviate many of the weaknesses that deteriorate past performance accuracy, enable



inadequate assessment justifications, and foster rater dissonance, while bolstering the government's ability to manage contractors' delivered performance levels.

In addition, several unsuspecting, novel factors emerged that explain past performance evaluation (in)accuracy. For example, informants attributed lower accuracy to evaluator turnover. This could be due to lower accountability for doing thorough work in observing and documenting contractor performance. Turnover can also exacerbate the problems caused by work overload. Thus, leaders should mitigate turnover of performance evaluators, particularly on complex contracts. Policy could also be enacted to require outgoing evaluators to conduct an interim CPAR prior to departure so that the new evaluator can begin—and assume accountability for assessing—performance evaluation at the beginning of a full evaluation period. Accuracy of past performance assessments was also affected by insufficiently defined requirements. It is difficult to assess that which is not understood or that which can have multiple interpretations. Thus, contracting officers and program managers should not move forward in contracting with ill-defined requirements. Additionally, contractors should strive to ensure that the buyer thoroughly defines requirements. An independent agency requirements ombudsman could help in this regard.

Perhaps the most novel finding is that the buyer's perceived fairness of the evaluation affects the accuracy of evaluations. This fairness can work for or against the contractor—depending on the buyer's assessment of what the contractor deserves. On the positive side, many informants likened the one-shot, summary rating that is supposed to reflect many instances of performance to an employee's annual performance appraisal. In other words, evaluators felt it unfair to rate a contractor as below satisfactory for a single instance of a performance failure in cases where there were many other performance opportunities. Similarly, performance evaluators were reluctant to give a below satisfactory rating singularly because of the impact to the contractor's ability to secure future government business. In addition to fear of a supplier dispute to ratings, this phenomenon confirms a halo effect. Conversely, on the negative side, some performance evaluators seemed to use the past performance rating as leverage—either as a threat to a contractor during performance and prior to a CPAR or as a means to punish a contractor following poor performance (i.e., revenge). The former was particularly acute involving contracts in which the government was locked in and had little relative bargaining power compared to that of the contractor (e.g., sole source contracts and those with high switching costs).

Theoretical Implications

Agency theory has been applied to many facets of buyer–supplier exchange relationships. In this study, two dimensions of agency operate simultaneously, and a third novel dimension emerged. First, the contractor is considered an agent of the buyer in promulgating the buyer's mission. Second, the buyer (i.e., the government team) is comprised of multiple agents to itself. In the case of multiple evaluators in different organizations of the government, multiple agency relationships exist, and each can hold different interests. The third unsuspected dimension of agency pertains to the program (i.e., the requirement). In some cases, both government performance evaluators and contractor employees could begin to identify more with the program than with their employer. In other words, sometimes, what is advantageous for the program can supersede what is advantageous for either the government or the contractor. This explains the halo effect afforded a contractor who fails in one instance of performance yet the government evaluator does not mention the failure in the past performance evaluation because of a reluctance to taint the program or the contractor's chance for future business. Thus, there appears to be opportunity to examine the antecedents and consequences of quasi-agency relationships to



understand under what circumstances such a quasi-agency emerges and the resultant effects.

Study Limitations

The obvious limitation of this paper is the lack of a quantitative test of the hypotheses. Thus, while serving as a foundation, future research should expand and test the propositions. These propositions lend themselves well to cross-sectional data collected via survey. The quantitative data could be analyzed using various multivariate models such as structural equation modeling. The research also employed a limited number of interviews. While rich insights were gleaned from experienced informants, other related phenomenon may be omitted with a narrow sample.

Future Research Directions

Future research should quantitatively test the hypotheses developed herein. Such a comprehensive model with many variables and successive dependent variables could be tested via structural equation modeling. Additionally, since the scope of this study was restricted to explaining past performance efficacy (i.e., its antecedents), the consequences of an effective past performance system should be empirically explored. In other words, does a more effective past performance system result in better source selection decisions, better contractor performance, and more efficient sourcing?

Future research could also expand the context of the study. This research was constrained to the federal government sector. Research could examine the extent to which the phenomenon occurs in the for-profit sector, and could examine differences in relationships among variables attributed to the differences in sectors. Hence, is the business sector a moderator for any of the hypothesized relationships?

Future research could also delve into situations in which performance evaluators empathize with the contractor to an extent that they are willing to inflate ratings and assessments. In other cases, we see just the opposite; performance evaluators are willing to use the past performance evaluation as a sort of punishment in a vengeful way. It would be interesting to understand why different evaluators in different situations take such different approaches.

References

- Beausoleil, J. W. (2010). *Past performance handbook: Applying commercial practices to federal procurement* (2nd ed.). Vienna, VA: Management Concepts.
- Contractor Performance Assessment Reporting System (CPARS) Metrics (n.d.). Retrieved May 19, 2011, from <http://www.cpars.csd.disa.mil/cparsfiles/cpars/metrics.htm>
- Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of Management Review*, 14(1), 57–74.
- Ellram, L. M. (1996). The use of the case study method in logistics research. *Journal of Business Logistics*, 17(2), 93.
- Federal Procurement Data System–Next Generation (FPDS–NG). (n.d.). Retrieved from https://www.fpds.gov/fpdsng_cms/index.php/en/
- Flint, D. J., Woodruff, R. B., & Gardial, S. F. (2002). Exploring the phenomenon of customers' desired value change in a business-to-business context. *Journal of Marketing*, 66(4), 102–117.
- GAO. (2001). *Trends and challenges in acquiring services* (GAO-01-753T). Washington, DC: Author.



- GAO. (2009). *Federal contractors: Better performance information needed to support agency contract award decisions* (GAO-09-1032). Washington, DC: Author.
- Hawkins, T. (2014). *Exploring the Efficacy of the Government's Current Use of Past Performance Information*. Monterey, CA: Naval Postgraduate School.
- Kelman, S. (2010). Three ways to revitalize the use of past-performance data. *Federal Computer Week*. Retrieved from <http://fcw.com/blogs/lectern/2010/05/steve-kelman-past-performance-source-selection.aspx>
- Office of Federal Procurement Policy (OFPP). (2011, January 21). *Improving contractor past performance assessments: Summary of the Office of Federal Procurement Policy's review, and strategies for improvement* [Memorandum]. Washington, DC: Office of Management and Budget.
- Patrick Van Ecke, G. S., Freund, W., Goeskjaer, M., & Ooms, B. (2006). *Legal study on unfair commercial practices within b2b e-markets: Final report* (European Commission Study ENTR/04/69). Retrieved from http://www.pedz.uni-mannheim.de/daten/edz-h/gddb/06/b2b_2006_ls.pdf
- Wathne, K. H., & Heide, J. B. (2000). Opportunism in interfirm relationships: Forms, outcomes, and solutions. *Journal of Marketing*, 64(4), 36–51.
- Williamson, O. E. (1975). *Markets and hierarchies: Analysis and antitrust implications*. New York, NY: The Free Press.
- Yin, R. K. (2009). *Case study research: Design and methods*. Los Angeles, CA: Sage Publications.

Acknowledgments

The author would like to thank the Acquisition Research Program for providing funding and editing support for this research.





ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL
555 DYER ROAD, INGERSOLL HALL
MONTEREY, CA 93943

www.acquisitionresearch.net



The Efficacy of the Government's Use of Past Performance Information: An Exploratory Study

11th Annual Acquisition Research Symposium

Tim Hawkins, Lt Col, USAF (Ret.), Ph.D.

15 May 2014



Issues in Past Performance Evaluation

Purpose & Research Question

Theoretical Frameworks

- Agency Theory
- Organizational Behavior
- Channel Communication

Methodology

Results

Implications





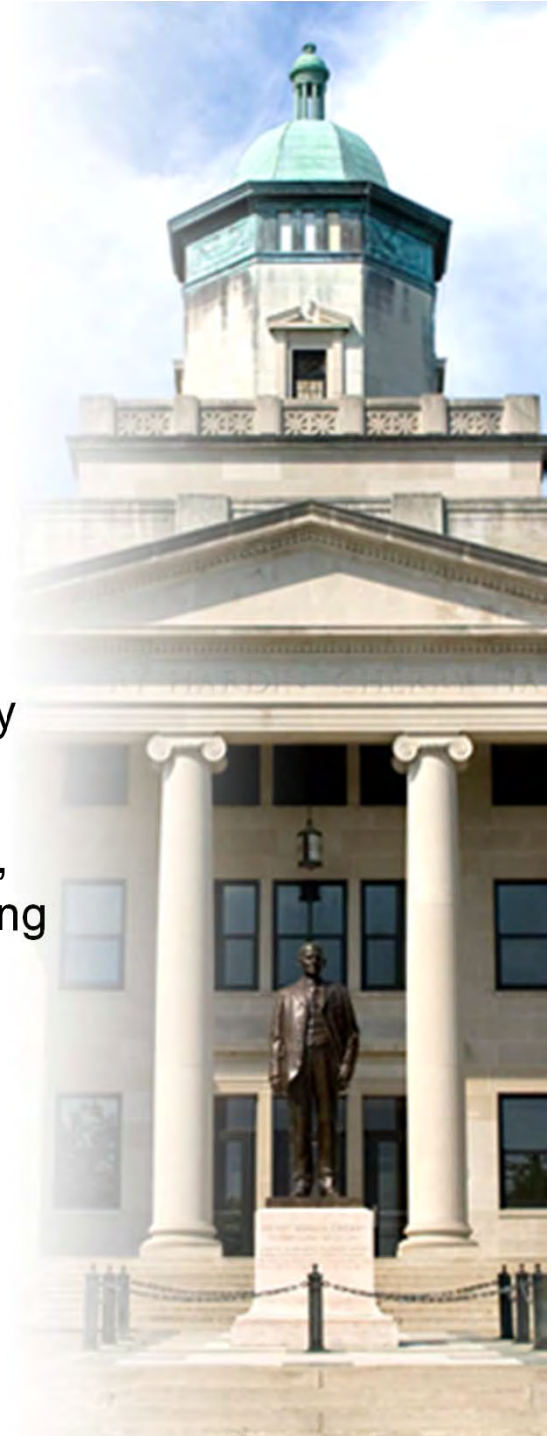
Issues in Past Performance Evaluation

- Only 31% of contract actions requiring CPARS reporting had completed reports (GAO,2009)
- Insufficient information to support ratings (OFPP, 2011)
 - how the contractor met, exceeded, or failed to meet requirements
- Incomplete reports – some categories not rated
- “*Halo Effect*” - Raters often inflate ratings to avoid conflict with the contractor (GAO, 2009)
 - PP increasingly subject to Contract Disputes Act
- Much attention and some improvement recently
 - Fed Gov’t PP Guide (2012), formerly DoD Guide (2011)



Problems

- Degree of inaccuracy of PPI unknown
 - Inaccurate PP assessments can harm contractors' reputations
 - Can bias source selections resulting in *adverse selection*.
- Reasons for inaccuracy not empirically explored
- Transaction costs not insignificant – but unknown precisely
 - If PPI is not reliable, and if evaluators cannot use the PPI to discriminate between proposals (Kelman, 2010), the effort of collecting and reporting, then later evaluating and documenting PPI is squandered
 - Federal contract managers are overworked (GAO, 2009) and understaffed (GAO, 2001)
 - Awarded 5.9M contract actions at \$538B in FY10
- PP evals, thus, often add little value to selection decisions





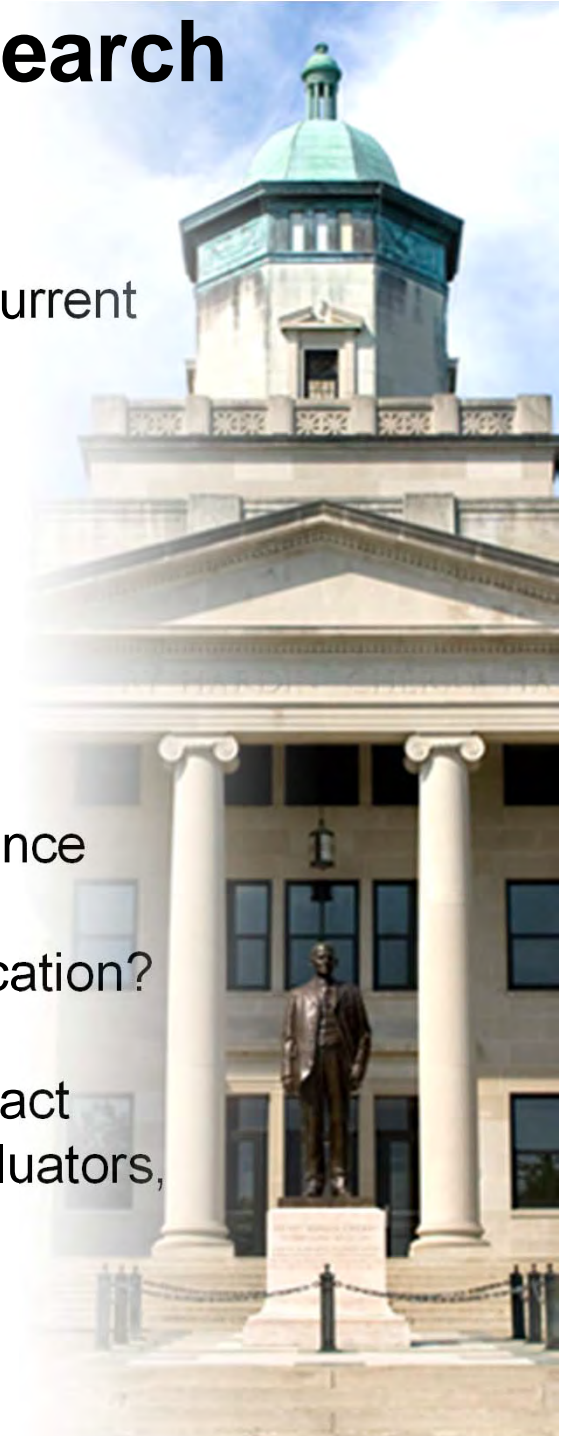
Purpose & Research Question

Purpose: Explore the efficacy of the government's current use of PPI

- Validate reported issues
- Tee up future research

Research Questions:

- Are PP reports useful?
 - Motivate suppliers to perform?
 - Reduce future performance uncertainty?
 - To what extent do PP evaluations/ratings influence source selection decisions?
- Why do PP evaluations/ratings lack sufficient justification?
- Why are PP evaluations sometimes inaccurate?
- In the cases of multiple evaluators on a single contract action, do PP evaluations/ratings deviate among evaluators, and, if so, why?
- Why do reviewing officials change the ratings of the evaluator (assessing official)?





Theoretical Frameworks

Agency Theory – 2 problems:

1. conflicting interests between principal and agent and
 2. difficulty and cost associated with monitoring agents, and the associated uncertainty for not having perfect information (Eisenhardt, 1989).
- Supplier as agent to buyer
 - Evaluator and other stakeholders as agents to buyer
 - Allegiance to buyer, program, or ktr (fairness; effect on ktr)?

Organizational Behavior

- PP likened to employee evaluations
 - Multiple raters
 - Halo effect

Channel Communication

- Formal Comm decr distortion (Mohr & Sohi, 1995)



Qualitative – appropriate when:

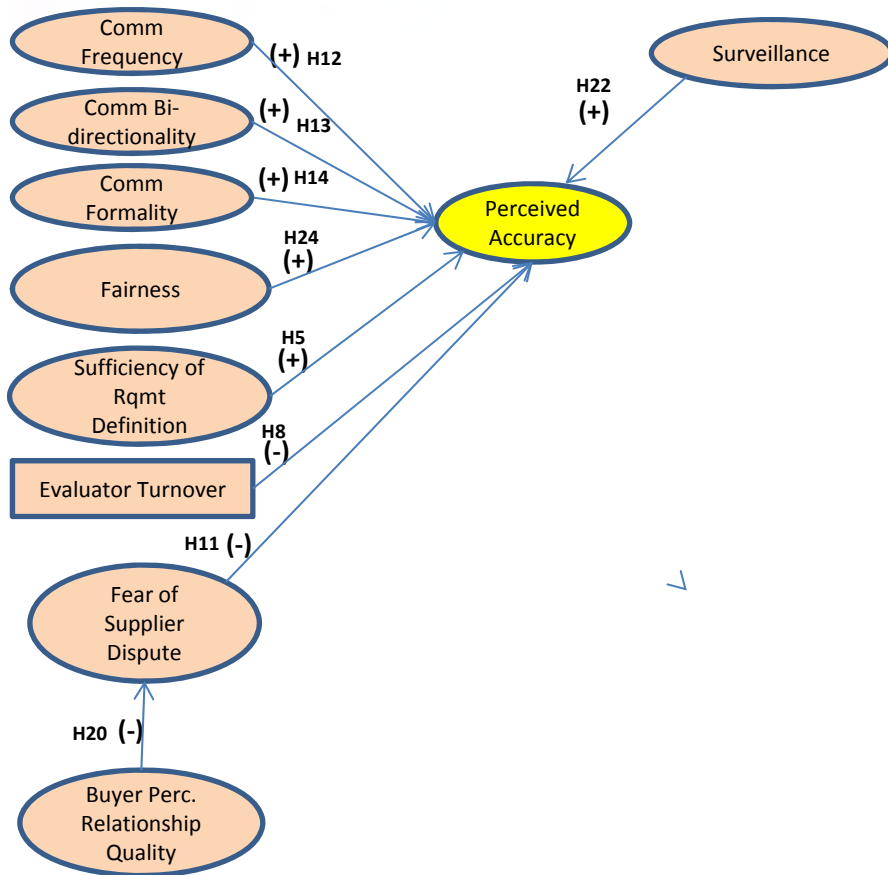
1. research is exploratory in nature (“why?”)
2. researcher has no control of the behavioral events being researched
3. focus is on contemporary events

Data Collection

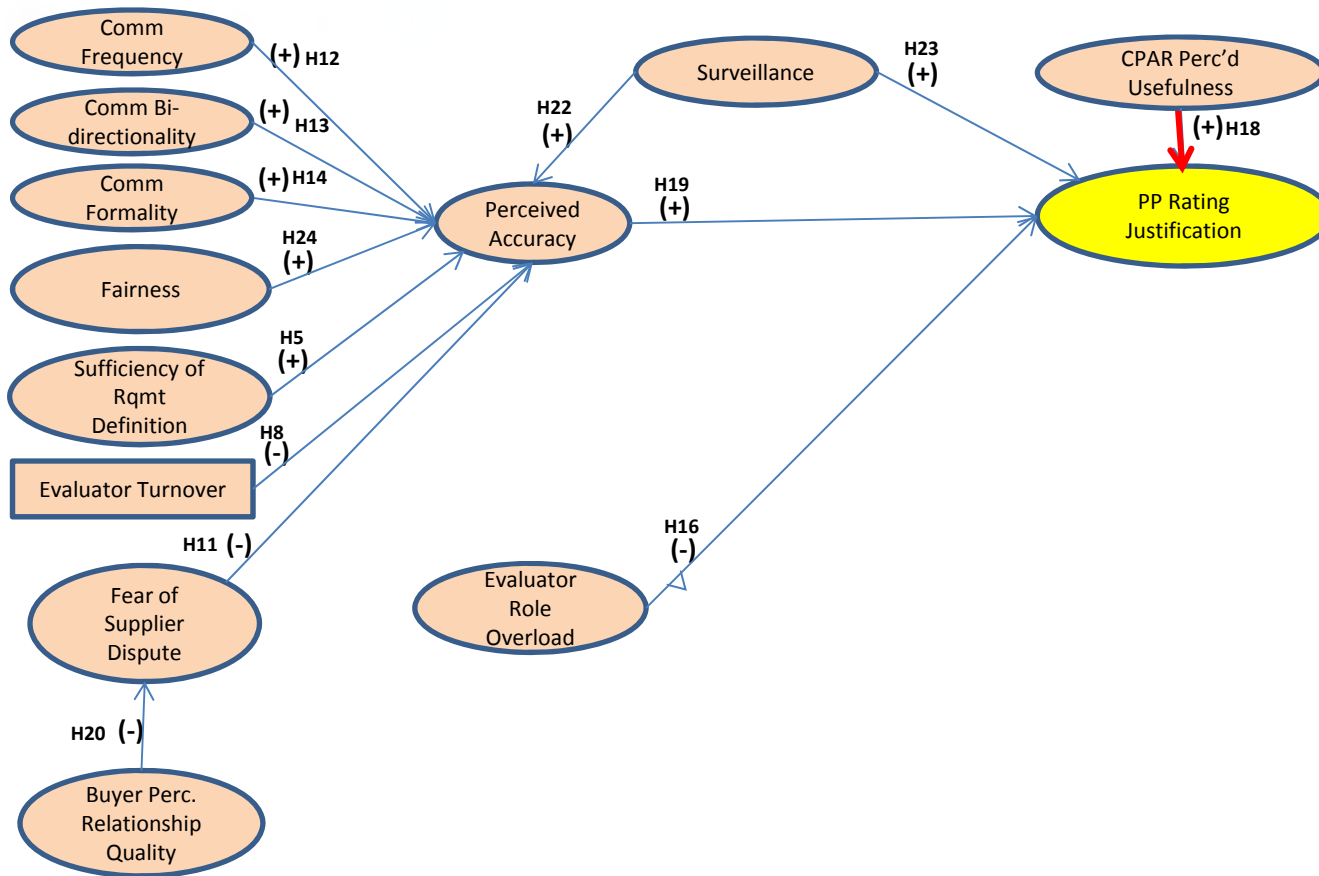
- Interview Protocol
- 8 Interviews
- 38-67 Min; avg 18 pages transcribed



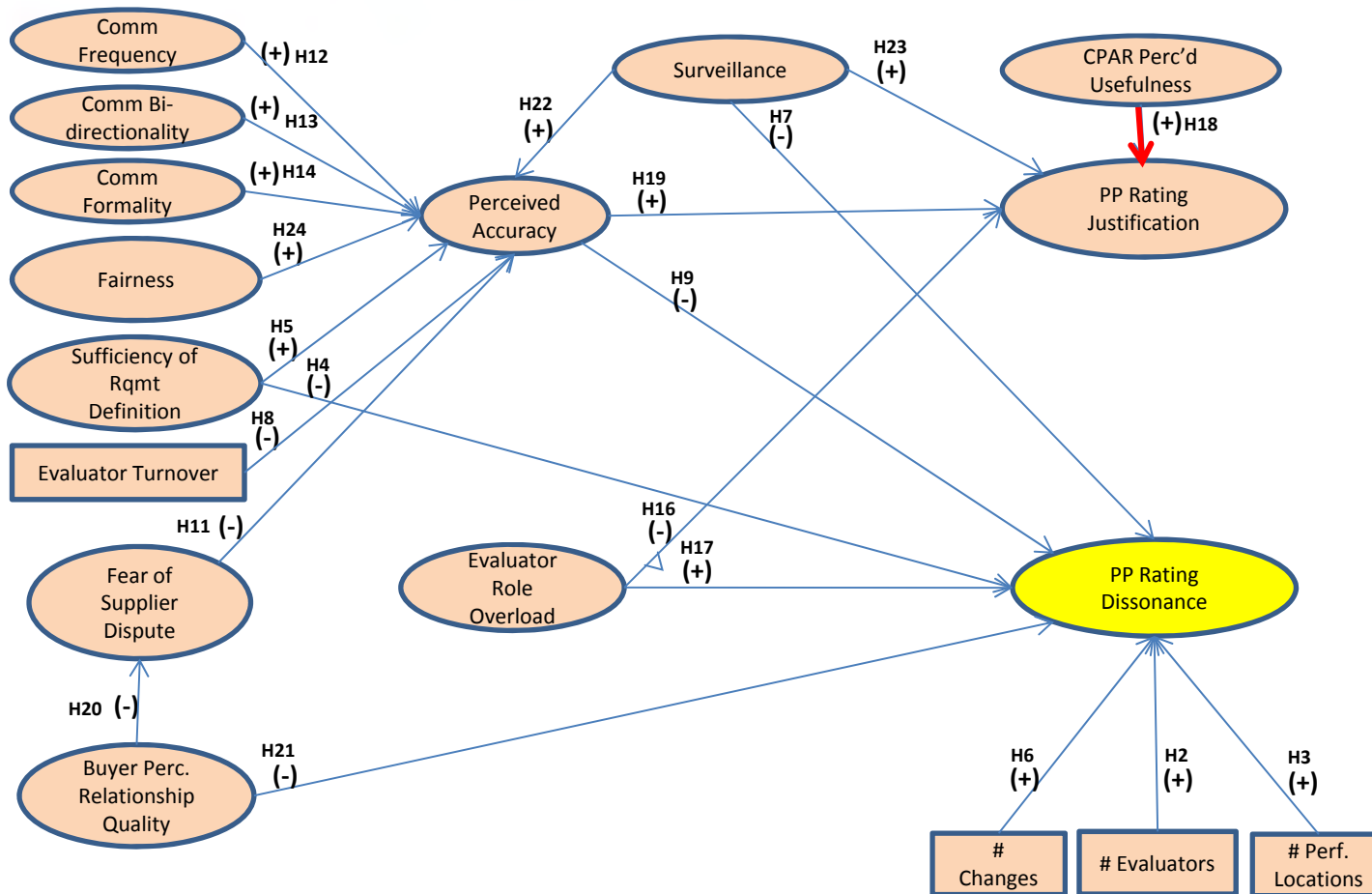
Results – Conceptual Model



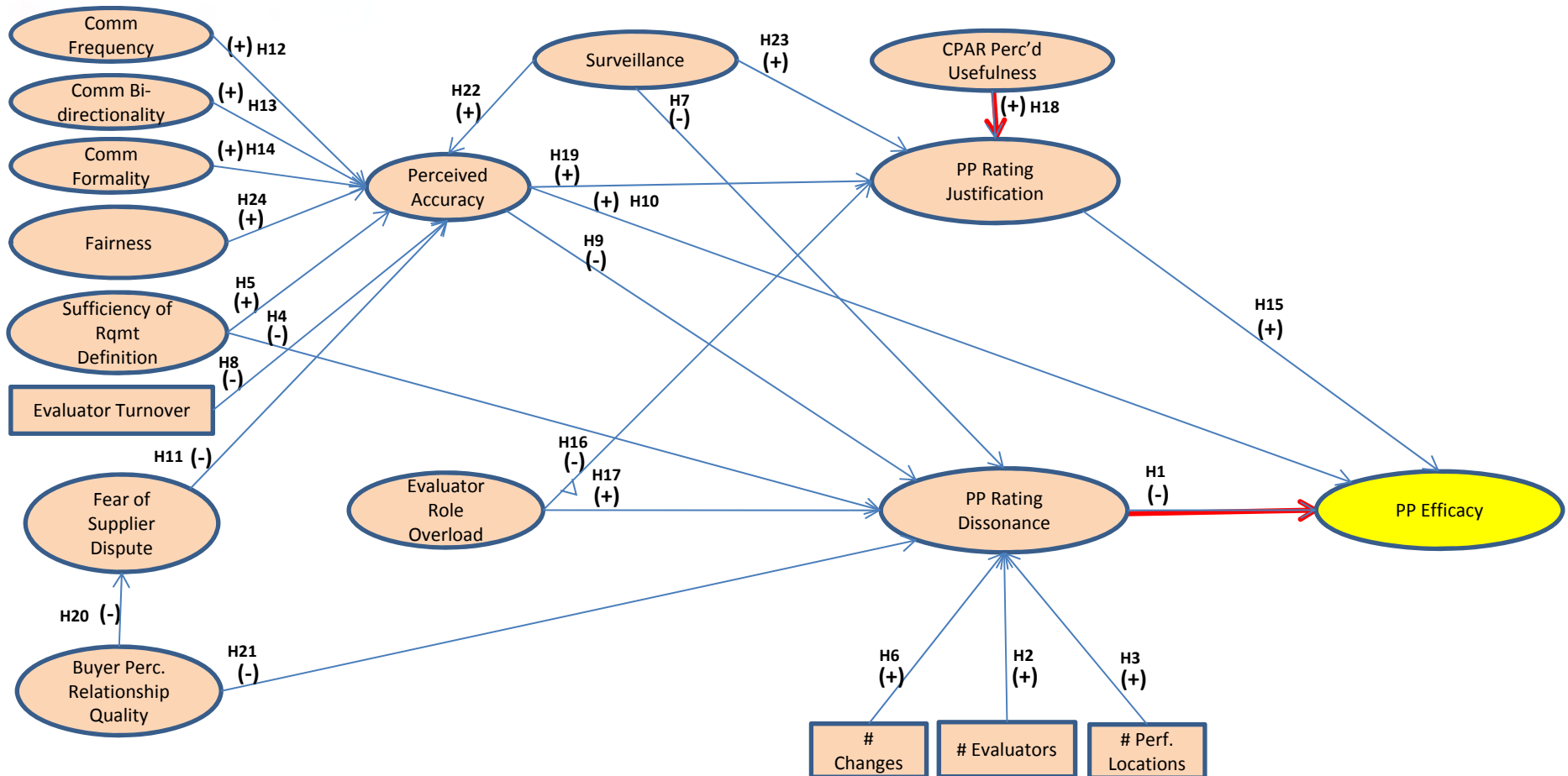
Results – Conceptual Model



Results – Conceptual Model



Results – Conceptual Model



Implications

- On average, past performance evaluations consume nearly 38 man-hrs of effort (rng 8-100).
- Leaders should ensure evaluators have sufficient time to perform their PP evals
 - Manning models need to account for PP workload
- Thoroughly define requirements—including measurements of success and precise definitions of CPAR ratings tailored to the requirement—prior to solicitation
- PP reporting is often not a sufficient surrogate for contractor performance management.
 - More frequent, formal, and two-way communication appears necessary to ensure rating accuracy



Implications (cont.)

- Leaders should mitigate turnover of performance evaluators
- Independent agency *requirements* ombudsman to ensure sufficient definitions of rqmts and PP ratings
- Halo effect confirmed – due to fear of dispute, fairness, protecting program, and concern for effect on ktr – particularly whether one instance of perf failure should represent all other successful perf opportunities
- Some use PP as leverage – as a threat ex ante, or as punishment ex post
- Lots of variance in performance info collection, recording, and sharing
 - Consider SPE system



Conclusion

- Confirmed many reported weaknesses
- Explained why the systemic weaknesses occur:
 - Accuracy of performance info,
 - Workload,
 - Variance in communications,
 - Poor rating justifications,
 - Variance in performance info collection, reporting, and sharing
- But, need to quantitatively confirm findings with a large sample
- Future research could explore effects of low PP efficacy on the contractor:
 - Performance?
 - Relationship quality?

